커뮤니티 별 Top 10 키워드 추출

1. 네이트판

['결혼', '남자', '남편', '본인', '부모', '시간', '엄마', '여자', '자기', '친구']

2. FM 코리아

['간부', '군대', '기부', '나라', '백신', '병원', '수술', '아웃', '추신수', '후원']

3. 루리웹

['매수', '민주당', '부모', '사건', '새끼', '유튜브', '의대', '자식', '친구', '한강']

4. 클리앙

['국가', '국민', '나라', '대만', '대통령', '백신', '일본', '중국', '처리', '한국']

프로토타입 감성 분석

1. 크롤링을 통해 각 커뮤니티 별로 데이터를 축적해 두고,

전체적으로 여러 키워드 추출 및 감성 분석 기법을 시도하여 개발

2. **각 커뮤니티 별**로 10개의 키워드를 추출하고,

이에 대한 감성 분석 결과를 도출

- 전날 데이터를 정리하여, 다음날 사용자에게 제공하고자 하므로,

하루 치 데이터를 추출하고 분석하는 방식으로 개발

전체 구조

키워드 추출 ~ 감성 분석 –예시

1. 선정 게시판 크롤링

fm.csv

2. 키워드 Top 10 추출

['간부', '군대', '기부', '나라', '백신', '병원', '수술', '아웃', '추신수', '후원']

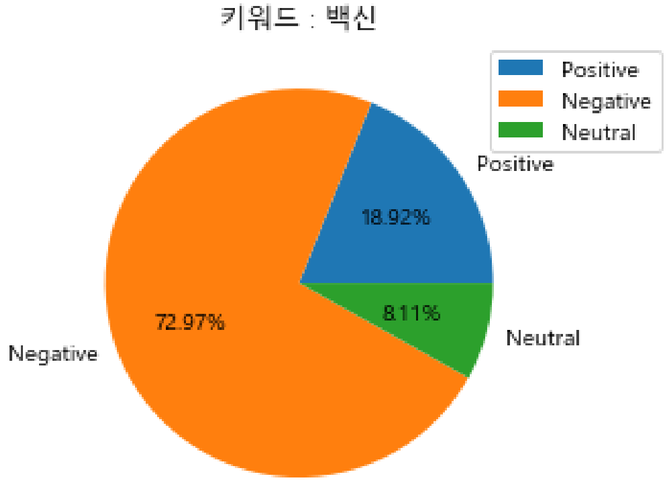
KNU 감성어 사전 활용하여 감성 분석

3. 각 키워드에 대한 데이터만 추출 및 토큰화

	index	제목	본문	댓글
0	5	오늘자 콜센터 상담사의 외 침	한남들 후원 취소하는 거 타격 도 없으니까 홈페이지로 취소 문의 넣어 달래	저런 거 보면 그냥 장작 넣을려고 위장 같음 드가 자재는 초록우산 측에서 백퍼 고소...
1	7	이 기적인 한남들 기부해봐야 얼마나 한다고	요즘 핫한 초록우산 페미 지원 뉴스베댓 한남들에게 일 침 남자 놈들 기부 해봐야 얼...	솔직히 말해서 여초애들이 후원을 하겠냐고 족보세트 하나라 더 시켜 먹으려고 할 건데...
2	9	초록우산 문의 답변 받음	내가 초록우산 한 년 넘게 후원하고 있는데 어제 초록우산 관련 글 보고 바로 문의 ...	그 디씨에 올라온 상담사 글이 일 수도 있겠네 한 남꺼져 이거구만 후원자가 하는 질...

4. 감성 분석

	Keyword	negative	neutral	positive
0	백신	-27	3	7

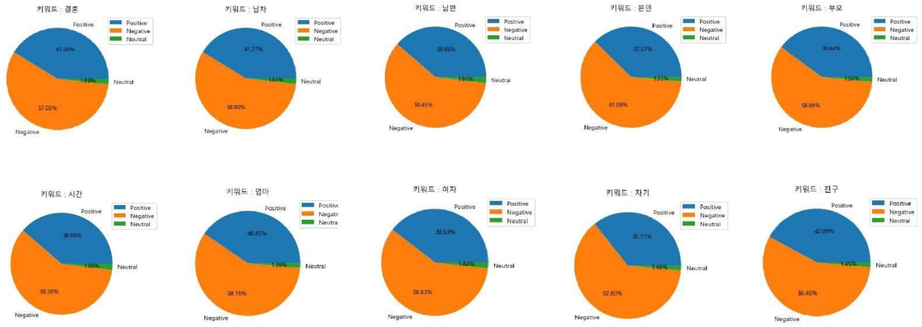


커뮤니티 별 감성 분석 결과

톡커들의 선택 명예의 전당

네이트판

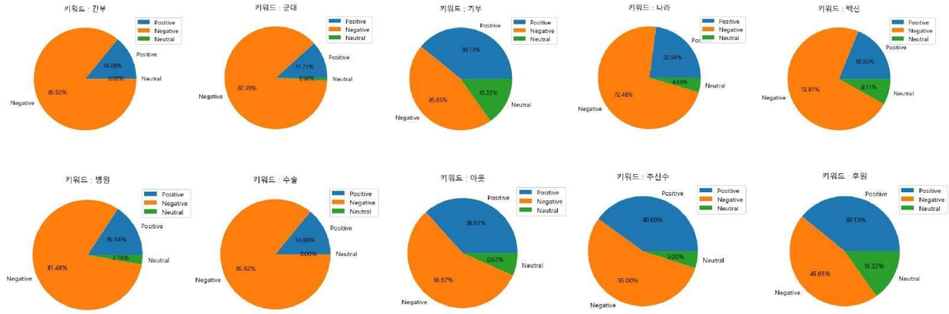
키워드: ['결혼', '남자', '남편', '본인', '부모', '시간', '엄마', '여자', '자기', '친구']



포텐 터짐 화제순

FM 코리아

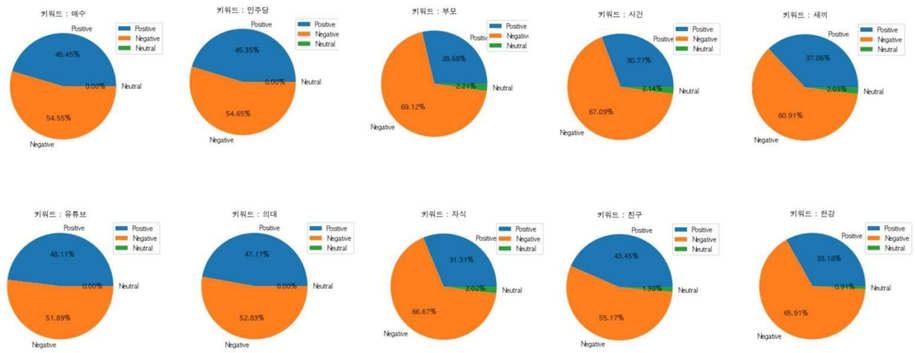
키워드: ['간부', '군대', '기부', '나라', '백신', '병원', '수술', '아웃', '추신수', '후원']



정치유머 게시판

루리웹

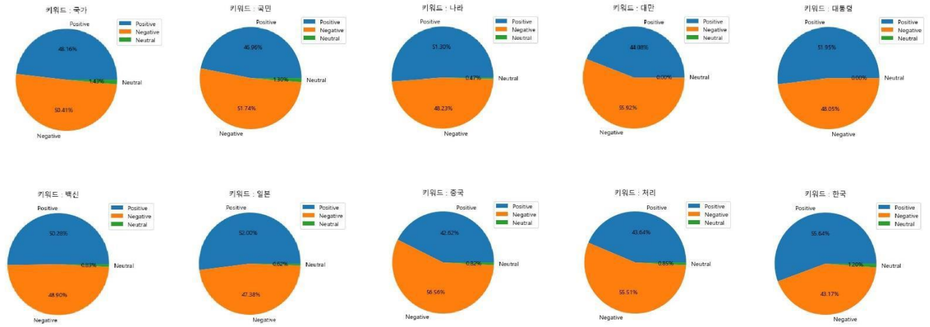
키워드: ['매수', '민주당', '부모', '사건', '새끼', '유튜브', '의대', '자식', '친구', '한강']



CLIEN 공감글

클리앙

키워드: ['국가', '국민', '나라', '대만', '대통령', '백신', '일본', '중국', '처리', '한국']



향후 계획

감성어 사전 추가 구축

SentiWord_Dict.txt

```
가볍게 행동하거나	-1
가볍게 행동하는	-1
가볍게 행동하다	-1
가볍고	1
가볍고 보드랍게	1
가볍고 상쾌하다	2
가볍고 상쾌한	2
가볍고 시원하게	2
가볍고 편안하게	2
가볍고 환하게	2
가분가분	1
가분히	1
가뿐가뿐	1
가뿐가뿐하다	1
가뿐가뿐히	1
가뿐하게	1
가뿐하다	1
가뿐한	1
가뿐한 느낌	1
가뿐한 느낌이	1
가뿐히	1
가쁘게	-1
가쁘게 쉬다	-1
가쁘고	-1
가쁘고 거칠게	-1
가쁘고 급하게	-1
가쁜	-1
가쁜 증상	-1
가살스럽다	-1
가소롭게	-1
```

KNU 한국어 감성어 사전에 없는 인터넷 용어나 난독화 된 단어들을 정리하고,

앞서 정한 Voting 기준에 따라 감성어를 추가하여,

분석의 품질, 즉 **정확도**와 **분석력**을 높이는 것을 목표로 하여 개발을 진행할 것이다.

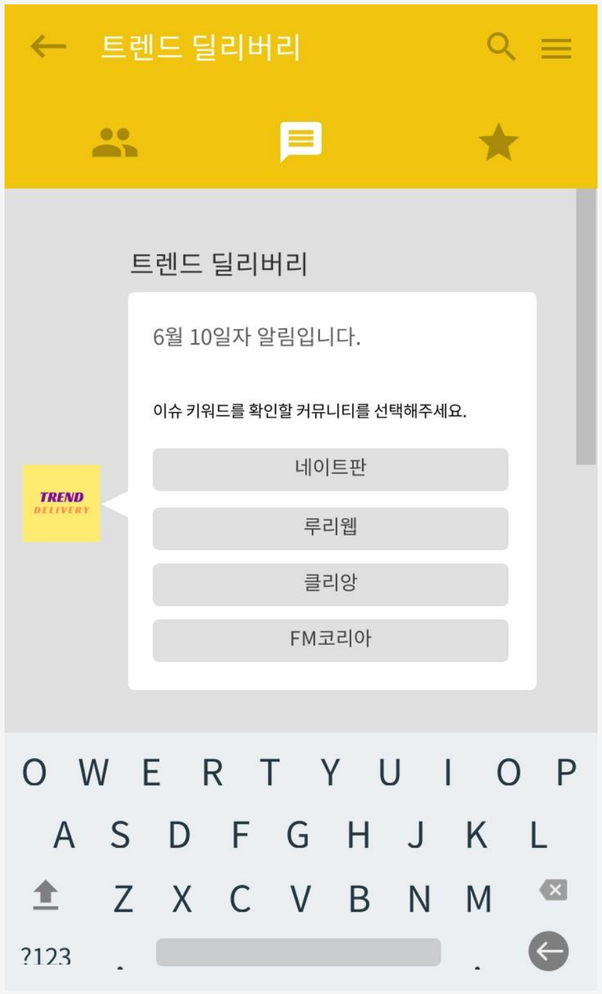
키워드 추출 방식 보완

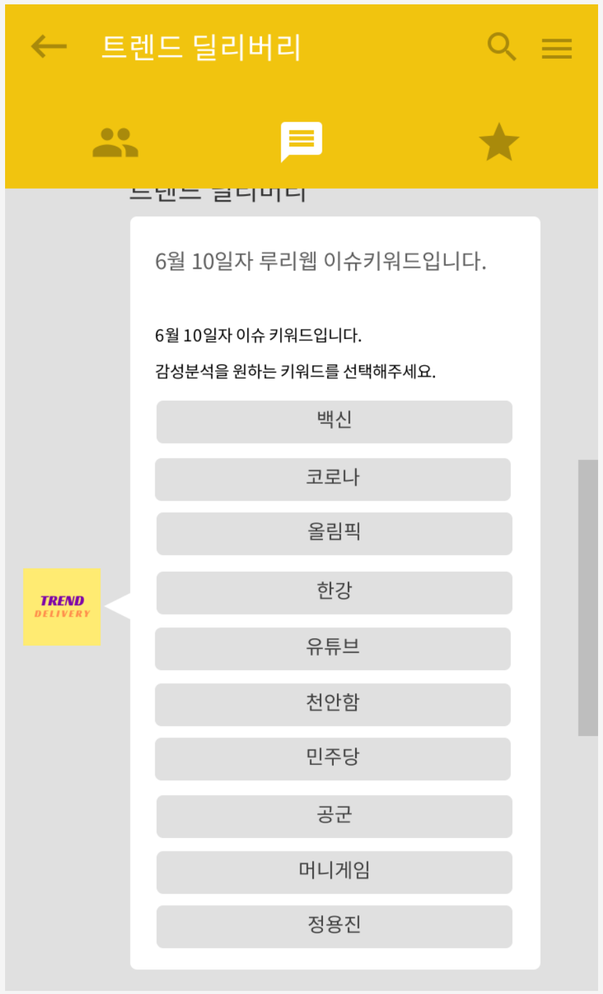
현재 프로토타입에서 사용 중인 TF-IDF 방식은 유사어 분류에 약하다는 단점이 있다.
이를 보완하기 위하여, 자주 혼용되는 아래의 기법들을 추가적으로 실습해보고 결과를 비교하여 사용할 예정이다. 이 과정을 통해 더 정확하고 유의미한 키워드를 추출하고자 한다.

- 토픽모델링의 일종인 LSA (Latent Semantic Analysis)
: 동음이의어 등 유사도 측정하여 의미론적으로 같은 내용을 묶어주는 방식

- Word2Vec
: 벡터 기법 중 하나인 일종의 얕은 신경망으로, 단어들 간의 유사성을 표현하는 방식

최종 목표: 챗봇 서비스





현재까지 제작된 프로토타입 모듈을 기초로 하여, 하나의 프로그램으로 웹페이지를 개발하고 관리자 시스템을 생성할 것이다. 사용자에게 챗봇 형식으로 매일 주기적인 시간에 알림을 제공하여, 전날의 이슈 키워드와 감성 분석 결과를 확인할 수 있게 해 주는 서비스를 제공할 계획이다.

개발 환경

개발 언어: python,

Back-End 프레임워크: django (장고, 파이썬 기반)

Front-End 프레임워크: BOOTSTRAP

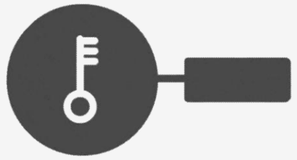
DB는 django 프레임워크에 기본적으로 지원되는 SQLite3를 사용합니다.

- 부트스트랩(Bootstrap)은 웹사이트를 쉽게 만들 수 있게 도와주는 프론트엔트 프레임워크로, 하나의 CSS로 휴대폰, 태블릿, 데스크탑까지 다양한 기기에서 작동하며 사용자가 쉽게 웹사이트를 제작, 유지, 보수할 수 있도록 기능을 지원하기 때문에 선택하였다.

- django는 파이썬으로 작성된 오픈 소스 웹 애플리케이션 프레임워크로, 쉽고 빠르게 웹사이트를 개발할 수 있는 구성요소로 이루어져 있다. 또한 파이썬 언어를 기반으로 하여 다양한 라이브러리들을 그대로 사용할 수 있다는 장점을 가지고 있기 때문에 채택하였다.

기대 효과

정치, 사회, 문화 등 다양한 분야에서의 온라인 여론 분석을 통해
체계적인 마케팅 전략을 수립할 정보 생성



01

매일 온라인 커뮤니티 이슈/여론을 간편하게 파악 가능

02

빅데이터 분석을 기반으로 추출한 키워드 및 감성 분석 데이터에서 인사이트 도출

03

도출한 인사이트를 바탕으로 맞춤형 마케팅 전략 수립 가능

감사합니다 :)

빅데이터미네이터

손윤식 교수님 주식회사 닥터송 송환구 멘토님